Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC).
A hyperspectral FT–IR-based approach to rapid screening for metabolite overproduction

Michael K. Winson[a], Royston Goodacre[a], Éadaoin M. Timmins[a], Alun Jones[a],
Bjørn K. Alsberg[a], Andrew M. Woodward[a], Jem J. Rowland[b], Douglas B. Kell[a,*]

[a]Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DA, UK
[b]Department of Computer Science, University of Wales, Aberystwyth, Ceredigion SY23 3DA, UK

Abstract

We introduce diffuse-reflectance absorbance spectroscopy in the mid-infrared as a novel method of chemical imaging for the rapid screening of biological samples for metabolite overproduction, using mixtures of ampicillin with *Escherichia coli* and *Staphylococcus aureus* as model systems. Deconvolution of the hyperspectral information provided by the raw diffuse reflectance-absorbance mid-infrared spectra was achieved using a combination of principal components analysis (PCA), artificial neural networks (ANNs) and partial least squares regression (PLS). Whereas a univariate approach necessitates appropriate data selection to remove any interferences, the chemometrics/hyperspectral approach could be employed to permit filtering of undesired components to give accurate quantification by PLS and ANNs without any preprocessing. The use of PCs as inputs to the ANNs decreased the training time from some 12 h to ca. 5 min. Equivalent concentrations of ampicillin between 0.05 and 20 mM in an *E. coli* or *S. aureus* background were quantified with >95% accuracy using this approach.

Keywords: Infrared spectroscopy; Multivariate calibration; Metabolic microscope; Strain improvement programmes; High throughput screening; DRASTIC; PLS; Neural networks

1. Introduction

As well as the increasing use of combinatorial chemical libraries [1–6], there is a large and continuing interest in the screening of microbial cultures for the production of biologically active metabolites (e.g. [7–20]), which can provide structural templates for synthetic programmes using rational methods of drug design. Methods based on synthetic oligonucleotides [21,22], phage display [23–25] and DNA shuffling [26,27] can provide further levels of diversity from biological starting points. Modern screens for such metabolites are targeted on the modulation of particular biochemical steps in the disease process and can

---

*Corresponding author. Tel.: +44 (1970) 622334; fax: +44 (1970) 622354; WWW: http://gepasi.dbs.aber.ac.uk/home.htm; e-mail: dbk@aber.ac.uk.

show a high degree of both specificity and sensitivity. This sensitivity means that metabolites showing activity during screening need be produced only in very small amounts by the organism. In such cases, increasing the titre of the metabolite is vital to provide enough material for further biological evaluation and chemical characterization and, eventually, for commercial production.

The process of titre improvement will normally involve the search for overproducing mutants derived from the original producing organism (see e.g. [12]), but titre-improving mutants are rare, typically at frequencies of 10^{-4} or less [28], and therefore many thousands of mutants need to be screened in search of an overproducing strain [29]. Previous methods of high-throughput screening (HTS) for mutants include the assessment of antibiotic activity of the metabolites (e.g. [30]) or use of rapid chromatographic methods such as thin layer chromatography (e.g. [31]) or fluorescence and luminescence methods such as the scintillation proximity assay [32–36]. Such methods historically can typically accommodate 10 000–50 000 isolates per month.

The ideal method for culture screening on plates (and indeed for the analysis of fermentor broths generally) would have minimum sample preparation, would analyse samples directly (i.e., be reagentless), would give information about recognizable chemical characters, and would be rapid, automated, non-invasive, quantitative and (at least relatively) inexpensive. These requirements indicate a spectroscopic solution, and we have recently demonstrated that the use of pyrolysis mass spectrometry (PyMS) in combination with a variety of chemometric methods allows rapid screening of cultures for metabolite overproduction [37–39], some 2 min per sample once these have been introduced to the carousel. However, the important conclusion to be drawn from these studies is that whole-cell or whole-broth spectral methods which measure all molecules simultaneously do contain enough spectral information from target molecules of interest to allow their quantification when the entire spectra are used as the inputs to modern chemometric methods based on supervised learning. The discriminatory power of these chemometric methods is such that in one study [37] we were able to assay quantitatively for the concentrations of ampicillin in mixtures with the Gram-positive *Staphylococcus aureus* when

the training set consisted of mixtures of ampicillin and an entirely different biological background, viz. the Gram-negative *Escherichia coli*. This shows, importantly, that chemometric methods of this type, which are designed to effect the quantification of biomolecules in complex biological backgrounds, may indeed be made highly resistant to changes in the background concentrations of metabolites and macromolecules.

As recently reviewed by Magee [40], the chemically based discrimination of intact microbial cells, referred to as whole-organism fingerprinting, involves the concurrent measurement of large number of spectral characters that together reflect the overall cell composition, the commonest spectral approach for this indeed being PyMS. There are, however, four general problems with using PyMS data as the input to supervised learning systems of this type: (i) the method is hardly non-destructive (although this is unimportant for broths, and for plates this could be dealt with by replica plating), (ii) it does not lend itself to in situ measurements, (iii) it still suffers somewhat from spectral drift (although recent advances suggest that this problem may be overcome [41]), and (iv) data acquisition still requires nearly 2 min per sample. Recently, a number of studies [42–47] have illustrated how even *visible* spectroscopy of petri plates could be used to identify colonies with high levels of electron transport chain components of interest, though this would not, of course, work directly for most target molecules. Most importantly, however, just as has been widely done with PyMS, Naumann and co-workers (e.g. [48–55]) have shown that FT–IR *absorbance* spectroscopy (in the mid-IR range, defined by IUPAC as 4000–200 cm^{-1}=2.5–50 µm) provides a powerful tool with sufficient resolving power to distinguish intact microbial cells at the strain level.

In view of the above, we therefore considered that the combination of FT–IR and *supervised* learning methods would permit us to extract the chemical concentration of the substance of interest, in a similar manner to that which we developed with PyMS. Sample preparation for absorbance measurements on biological samples of this type is rather tedious, however. Instead, and because FT–IR may be carried out using *reflectance* methods, we considered that one should seek to obtain spectra as a function of spatial

location, and by combining the spectroscopy with supervised learning methods obtain images in which metabolite concentrations are encoded as colours or contours, i.e. to construct a *metabolic microscope*. In this regard, it is particularly noteworthy that White's group [56,57] have shown the ability of diffuse reflectance FT–IR (DRIFT) spectroscopy without any chemometric processing, to effect the discrimination of microbes on surfaces. In a related vein, Yan and co-workers [58] recently showed that FT–IR could be used to analyse solid-state pins as used in combinatorial chemistry, whilst Gremlich and Berets [59] used FT–IR internal reflection spectroscopy for a similar purpose.

We therefore here describe the utilisation of our development of diffuse reflectance/absorbance FT–IR spectroscopy in the mid infra-red as a quantitative tool for the rapid analysis of all samples of biotechnological interest, specifically by exploiting the ability of modern, supervised learning methods to take multivariate spectral inputs and map them directly to the concentration of one or more target determinands (see above and [60]), using as before [37] mixtures of ampicillin with *E. coli* and *S. aureus* as model systems.

2. Experimental

2.1. Preparation of mixtures of ampicillin with E. coli and S. aureus

The bacterial strains used were *E. coli* HB101 [61] and *S. aureus* NTCC 6511. Both strains are ampicillin-sensitive, indicating that any spectral features observed are not due, for instance, to β-lactamase activity. The mixtures were prepared as previously [37]. The strains were grown in 500 ml liquid medium (glucose (BDH), 10.0 g; peptone (LabM), 5.0 g; beef extract (LabM), 3.0 g; per litre water) for 16 h at 37°C in a shaker. After growth the cultures were harvested by centrifugation, washed and resuspended in physiological saline (0.9% NaCl). Ampicillin (desiccated D[-]-α-aminobenzylpenicillin sodium salt, $\leq$98% (titration), Sigma) was prepared to give final concentration ranges of 0.05–2 mM in 0.05 mM steps and 0.5–20 mM in 0.5 mM steps in suspensions of 18 mg ml^{-1} (dry weight) of bacterial cells.

2.2. Diffuse reflectance–absorbance FT–IR

Four replicate 5 µl aliquots of the above samples were transferred into wells in a sandblasted aluminium plate (measuring 10 cm by 10 cm) and dried at 50°C for 30 min. The plate was mounted onto a motorised stage and the samples analysed using a diffuse reflectance TLC accessory [62–64] connected to a Bruker IFS28 FT–IR spectrometer (Bruker Spectrospin Ltd., Banner Lane, Coventry CV4 9GH, UK) equipped with a liquid N_2-cooled MCT (mercury-cadmium-telluride) detector. A schematic of the general optical arrangement of this accessory is shown in Fig. 1.

The IBM-compatible PC used to control the IFS28 was programmed (using OPUS version 2.1 software running under IBM OS/2 Warp provided by the manufacturers) to collect spectra over the wavenumber range 4000–600 cm^{-1}. Spectra were acquired at a rate of 20 s^{-1} and the spectral resolution was 4 cm^{-1}. To improve the signal-to-noise ratio 256 spectra were co-added and averaged. Each sample was represented by a spectrum containing 882 points, and spectra were displayed in terms of absorbance using the Opus software.
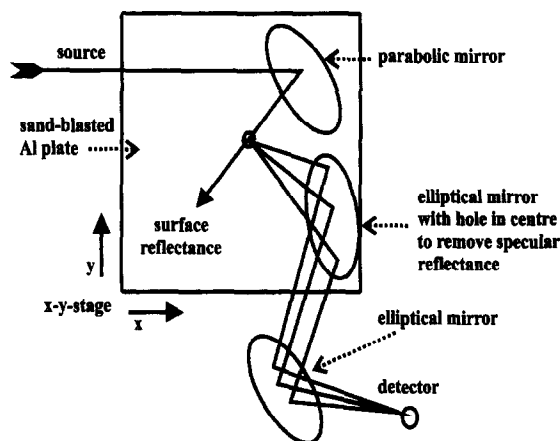

Fig. 1. Scheme of the Bruker TLC unit for diffuse reflectance–absorbance Fourier transform infrared spectroscopy. Infrared light passes through the sample, is diffusely reflected from the rough metal backing plate, passes through the sample again and is sent to the detector. The elliptical mirror contains a small hole to remove light reflected specularly.

2.3. Pre-processing and exploratory analysis

Spectral data were exported from the Opus software used to control the FT–IR instrument, converted to ASCII format and imported into Matlab version 4.2c.1 (The MathWorks, Inc., 24 Prime Park Way, Natick, MA, USA), which runs under Microsoft Windows NT on an IBM-compatible PC.

2.4. PCA and PLS

Matlab was used to perform principal components analysis (PCA) according to the NIPALS algorithm [65], so that exploratory data analysis could be conducted. PCA is a multivariate statistical technique which can be used to identify *correlations* amongst a set of variables (in this case 882 wavenumbers) and to transform the original set of variables to a new set of *uncorrelated* variables called principal components (PCs). The objective of PCA is to see if the first few PCs account for most (>90%) of the variation in the original data [66]. If they do reduce the number of dimensions required to display the observed relationships, then the PCs can more easily be plotted and 'clusters' in the data visualized [67]; moreover this technique can be used to detect outliers [68]. PLS modelling was performed as previously described [69]. PLS is a multivariate technique similar to PCA, but with the components extracted using both *x*- and *y*-data and then regressed onto the (known) training results while forming the model. This results in a more parsimonious model in situations where the variance of interest may not be the largest variance in the samples.

2.5. Artificial neural networks

All artificial neural network (ANN) analyses were carried out with a user-friendly, neural network simulation program, NeuFrame version 1,1,0,0 (Neural Computer Sciences, Lulworth Business Centre, Nutwood Way, Totton, Southampton, Hants), which runs under Microsoft Windows NT on an IBM-compatible PC. In-depth descriptions of the modus operandi of this type of ANN analysis are given elsewhere [37–39,41,64,70–80].

Before training commenced, the values applied to the input and output nodes were normalised between 0 and 1, and the connection weights set to small random values [81]. The network was trained by the standard back-propagation method [82,83]. Each training epoch represented connection weight updatings and a recalculation of the root mean squared (RMS) error between the true and desired outputs over the entire training set. This process was repeated until an acceptable level of error was achieved.

3. Results and discussion

FT–IR analysis of ampicillin in mixtures with bacterial cells provides a model for determining the presence of a metabolite of interest in industrial fermentations. Many studies on the quantification of particular determinands in mixtures using FT–IR have been based on the contribution of only one or a few spectral features. The FT–IR spectrum of ampicillin contains such a characteristic marker band at $\sim$1767 cm^{-1} [84] which is lost after cleavage of the β-lactam ring by β-lactamase (Fig. 2). Integration of the peak area at 1767 cm^{-1} from diffuse reflectance FT–IR measurements for a range of concentrations (0.5–20 mM) of ampicillin mixed with *S. aureus* was used in a simple (quasi-)univariate linear regression (i.e., $y = mx + c$) in order to estimate the ampicillin
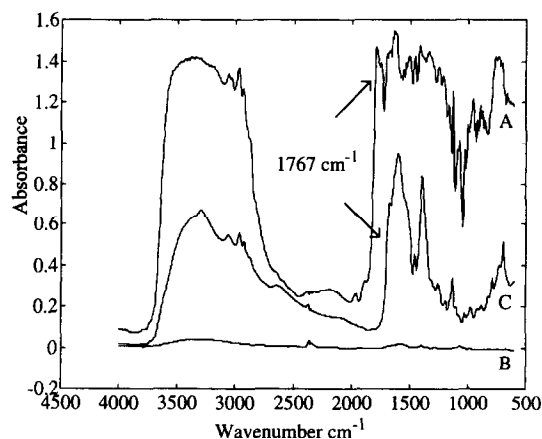

Fig. 2. Typical FT–IR diffuse reflectance-absorbance spectra of (A) 100 μmol ml^{-1} ampicillin (desiccated D[-]-α-aminobenzylpenicillin sodium salt, ≥98% (titration), Sigma); (B) 1000 units ml^{-1} β-lactamase (Type 1 from *Bacillus cereus*. One unit will hydrolyse 1.0 μmole of ampicillin per min at pH 7.0 at 25°C, Sigma) and (C) 'A' following the addition of 20 units of β-lactamase in 'B' and incubation at 25°C for 20 min.

titre. The 20 integrals in the training set were fitted linearly to the concentrations to generate the expression $y = 0.16x + 0.827$. This expression was then used to predict the training and test sets, producing RMS errors of 1.27 and 1.21 respectively. Although in this particular case a reasonable degree of accuracy was obtained by using simple linear regression analysis on a single spectral feature, it is likely that for most analyses of authentic fermentations this type of single-band method could not be successfully applied because of the cell background variability. Multivariate analysis, however, in combination with modern chemometric techniques such as PLS and ANNs, allows us to predict the ampicillin concentration using full-spectrum calibration. This approach was used to predict the concentration of ampicillin in mixtures with *E. coli* and *S. aureus* cells from the full spectral results using PLS regression [69,85–87], artificial neural networks (ANNs) and PCA scores as input to ANNs (PC-ANNs).

Fig. 3 shows the FT–IR spectra for a range of concentrations of ampicillin in mixtures with *S. aureus* and *E. coli*. Initially, all 882 data points derived from the absorbance spectra were used as inputs for training ANNs. Fig. 4 shows the result of using neural networks to predict the concentration of ampicillin in the range 0.5–20 mM in mixtures with *S. aureus* cells using this multivariate method. However, because of the large number of inputs this approach proved to be particularly time consuming (and typically took 12 h to compute), and at all events the use of all the variables means that some will be irrelevant or collinear and thus detrimental to the quality of the model formed [88–91]. To obey the so-called Parsimony principle [90] and to circumvent the possibility of over-fitting, PCA was used to reduce the number of input nodes [64]. Fig. 5 shows the effect of using PCs on the ability of X-4-1 ANNs (where X is the number of PCs applied to the input nodes) to quantify the levels of ampicillin in an *E. coli* background. When very few PCs are used (<4 PCs) generalisation is poor because insufficient spectral information is accounted for; conversely as the number of PCs used is increased these incorporate noise and the predictive ability of the ANNs decrease. All six mixture data sets were reduced using PCA and the PCs used to train ANNs. An example of the improvement in the accuracy of prediction for quantification of ampicillin in a mixture
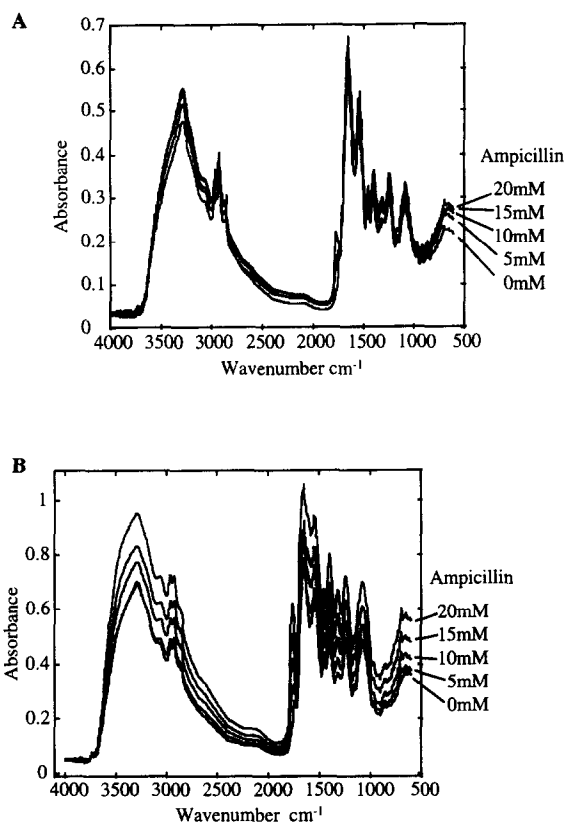

Fig. 3. FT–IR diffuse reflectance–absorbance spectra of (A) *E. coli*; and (B) *S. aureus*; with different concentrations of ampicillin.

with *S. aureus* (for the 0.5–20 mM ampicillin range) is shown in the comparison of the results for a PC-ANN trained with 15 PCs vs. one trained on the total spectra (882 data point); the RMS errors for the test sets were 0.503 and 0.874 respectively (Figs. 4 and 6 and Table 1) whilst the number of epochs was 900 for the reduced data and 50 000 for the full spectral approach; this equates to a decrease in training time of from approximately 12 h to only 5 min, whilst also improving the ANNs predictive ability.

The results from the multivariate analysis using PLS, ANNs and PC-ANNs clearly shows that it is possible to form a model capable of quantifying unknown concentrations of ampicillin between 0.05 mM and 20 mM from either an *E. coli* or *S. aureus* cell background. In order to see if the background (i.e., the *E. coli* or *S. aureus* cells) affected the ability to determine the concentration of ampicillin, a
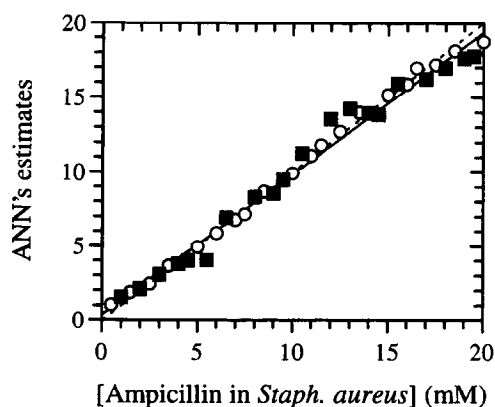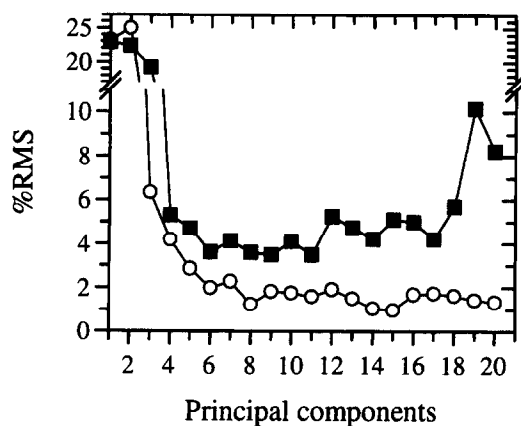
O Training set   · · · · Expected proportional fit

■ Test set   —— Calculated linear fit on test set

Fig. 4. The estimates of trained 882-10-1 neural networks vs. the true ampicillin concentration (0.5–20 mM) in S. aureus. The input layer was scaled for each input node such that the lowest absorbance for each wavenumber was set to 0 and the highest to 1. The networks were trained using the standard back propagation algorithm for approximately 5×10^4 epochs. Open circles represent spectra that were used to train the network and closed squares indicate 'unknown' spectra which were not in the training set. The calculated linear fit (bold line) and expected proportional fit (broken line) are shown. The RMSEP for the test set was 0.874.

combined model based on spectral data from both sets of backgrounds was used for predicting unknown concentrations from spectral data of ampicillin in either an *E. coli* or *S. aureus* background alone (Table 2). Use of the combined model for predicting the ampicillin concentration (in the range 0.5–20 mM) with PC-ANNs produced a test set RMS value of 0.75 (Table 2) compared with 0.611 for *E. coli* and 0.503 for *S. aureus* models used with their cognate sets (Tables 1 and 3).

Diffuse reflectance FT–IR can be used to produce maps of spectral profiles from surface scans. After processing of the data by chemometric methods, an image (consisting of a colour or surface contour map) of a particular metabolite(s) can be constructed. We were able to utilise this approach (using full spectrum calibration by PLS) to build up a chemical image map showing the concentrations and distribution of ampicillin applied to the surface of a metal plate in an *E. coli* cell background (Fig. 7).

-O- Training set   ■ Test set

Fig. 5. An example of the effect on % RMSEC and % RMSEP of training X-4-1 ANNs with test set cross-validation, where X is the number of PC scores used to train the ANNs. Open circles represent spectra that were used to train the network and closed squares indicate 'unknown' spectra which were not in the training set.
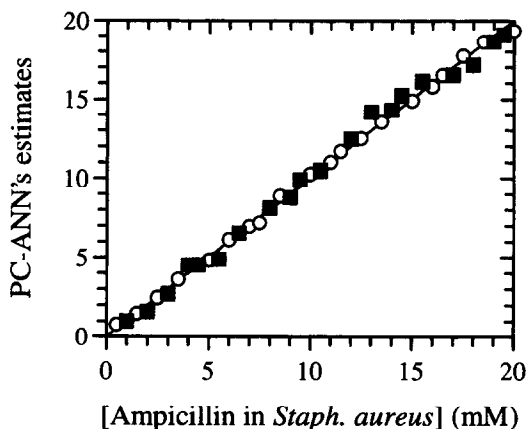
O Training set   · · · · Expected linear fit

■ Test set   —— Calculated best fit on test set

Fig. 6. The estimates of trained 15-4-1 neural networks vs. the true ampicillin concentration (0.5–20 mM) in S. aureus. The input layer was scaled for each input node such that the lowest PC score was set to 0 and the highest PC score to 1. The networks were trained using the standard back propagation algorithm for 900 epochs. Open circles represent spectra that were used to train the network and closed squares indicate 'unknown' spectra which were not in the training set. The calculated linear fit (bold line) and expected proportional fit (broken line) are shown. The RMSEP for the test set was 0.503.

Table 1

Comparison of partial least squares (PLS), artificial neural network (ANNs) and ANNs with principal component scores as input (PC-ANNs) in the deconvolution of diffuse reflectance FT-IR spectra from ampicillin mixed in S. aureus

Method	0.05–2 mM Ampicillin			0.5–20 mM Ampicillin		
	PLS	ANNs	PC-ANNs	PLS	ANNs	PC-ANNs
Factors	7			8		
Epochs		50 000	1000		50 000	900
PCs			15			15
RMSEP						
Training set	0.061	0.066	0.048	0.294	0.398	0.229
Test set	0.066	0.074	0.064	0.540	0.874	0.503

Table 2

Comparison of partial least squares (PLS), artificial neural network (ANNs) and ANNs with principal components scores as input (PC-ANNs), for predicting the concentration of ampicillin in a test set containing ampicillin mixed in either E. coli or in S. aureus using a model trained on a combination of diffuse reflectance FT–IR spectra from cognate sets

Method	0.05–2 mM Ampicillin			0.5–20 mM Ampicillin		
	PLS	ANNs	PC-ANNs	PLS	ANNs	PC-ANNs
Factors	10			9		
Epochs		100 000	6000		100 000	2250
PCs			20			15
RMSEP						
Training set	0.114	0.050	0.046	1.484	0.664	0.669
Test set	0.138	0.068	0.072	1.206	0.350	0.750

Table 3

Comparison of partial least squares (PLS), artificial neural network (ANNs) and ANN with principal component scores as input (PC-ANNs) in the deconvolution of diffuse reflectance–absorbance FT–IR spectra from ampicillin mixed in E. coli

Method	0.05–2 mM Ampicillin			0.5–20 mM Ampicillin		
	PLS	ANNs	PC–ANNs	PLS	ANNs	PC–ANNs
Factors	7			6		
Epochs		50 000	15 000		50 000	3000
PC's			15			20
RMSEP						
Training set	0.057	0.038	0.034	0.898	0.482	0.415
Test set	0.1	0.093	0.075	0.743	0.879	0.611

4. Conclusions

Driven in part by the activities of the remote sensing community [92], there is a burgeoning interest in the rapid acquisition of diffuse reflectance spectral data from various spatial locations, detecting hundreds of wavelengths simultaneously (most commonly in the visible and near infrared), and coupled increasingly to advanced data reduction and visualization algorithms, an approach often referred to as hyperspectral imaging [93–101]. Such remote-sensing analyses occasionally use the mid-IR part of the spectrum [102–105]. One of the problems with this approach to *remote* sensing, however, is the strong and variable absorbance of radiation by the atmosphere itself [101,106,107], a problem from which we do not suffer.
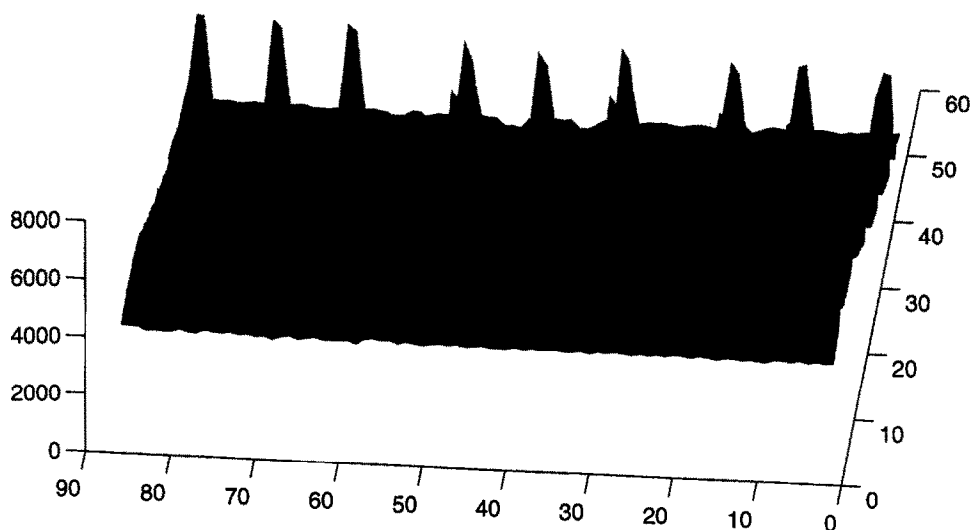
Fig. 7. A chemical image derived from the FT–IR spectra of samples of *E. coli* with different concentrations of ampicillin (μg ml^{-1}) on the surface of a sandblasted aluminium plate. FT–IR spectra were recorded at 1 mm intervals and the peak values were based on full spectrum calibration using PLS.

Diffuse reflectance FTIR in combination with a multivariate calibration chemometric approach to data analysis could be used to effect the rapid quantification of a pharmaceutical product (ampicillin) in a (variable) biological background (*E. coli* and *S. aureus* cells), a situation representative of metabolite overproduction in a screening or titre improvement programme [37]. Spectral variation contributed by shifting baseline due to instrumental interference and differences in the biological background between samples could largely be eliminated by PC-ANNs, thus giving an RMS error of 0.503 compared with 1.21 from using the 'quasi'-univariate approach for 0.5–20 mM ampicillin in *S. aureus*. Although diffuse reflectance methods are well known to suffer difficulties in quantitative work (the concentration region for which Kubelka–Munk theory [108–111] holds is normally quite small, for instance [112]), it is clear that the combination of modern chemometric methods with the diffuse reflectance–absorbance approach overcomes these most satisfactorily. As an analytical tool for HTS the DRASTIC approach (using a spectrometer with a TLC accessory and automated sample handling) would be capable of quantifying a particular determinand from in excess of 50 000 samples *per day*, a substantial improvement on current processing capability using traditional methods such as HPLC. Thus

we have here shown for the first time that the hyperspectral approach using diffuse reflectance–absorbance spectroscopy coupled to modern supervised learning methods provides a novel, rapid, powerful and general approach to the problem of screening for metabolite overproduction in biological and biotechnological systems.

Acknowledgements

References

[1] N.K. Terrett, M. Gardner, D.W. Gordon, R.J. Kobylecki and J. Steele, Tetrahedron, 51 (1995) 8135–8173.
[2] M.A. Gallop, R.W. Barrett, W.J. Dower, S.P.A. Fodor and E.M. Gordon, J. Med. Chem., 37 (1994) 1233–1251.
[3] E.M. Gordon, R.W. Barrett, W.J. Dower, S.P.A. Fodor and M.A. Gallop, J. Med. Chem., 37 (1994) 1385–1401.

[4] S.E. Blondelle, E. Perezpaya, C.T. Dooley, C. Pinilla and R.A. Houghten, Trac-Trends Anal. Chem., 14 (1995) 83–92.

[5] D.J. Ecker and S.T. Crooke, Bio/Technology, 13 (1995) 351–360.

[6] K.D. Janda, Proc. Natl. Acad. Sci., 23 (1994) 10779–10785.

[7] J.D. Bu'Lock, L.J. Nisbet and D.J. Winstanley, Bioactive microbial products: Search and discovery, Academic, New York, 1982.

[8] R.P. Elander, in J. Bu'lock and B. Kristiansen, (Eds.) Basic Biotechnology, Academic Press, London, 1987, pp. 217–251.

[9] E.J. Vandamme, Biotechnology in Industrial Antibiotics, Marcel Dekker, New York, 1984.

[10] M.S. Verrall, Discovery and Isolation of Microbial Products, Ellis Horwood, Chichester, 1985.

[11] S. Omura, Microbiol. Rev., 50 (1986) 259–279.

[12] W. Crueger and A. Crueger, Biotechnology: A Textbook of Industrial Microbiology, Sinauer Associates, Sunderland, MA, 1989.

[13] N. Porter and F. Fox, Pesticide Sci., 39 (1993) 161–168.

[14] L.J. Nisbet and N. Porter, Symp. Soc. Gen. Microbiol., 44 (1989) 309–342.

[15] P.D. Crittenden and N. Porter, Trends Biotechnol., 9 (1991) 409–414.

[16] C.M.M. Franco and L.E.L. Countinho, CRC Crit. Rev. Biotechnol., 11 (1991) 193–276.

[17] G. Lancini and R. Lorenzetti, Biotechnology of antibiotics and other bioactive microbial metabolites, Plenum Press, New York, 1993.

[18] P. Bevan, H. Ryder and A. Shaw, Trends Biotechnol., 13 (1995) 115–121.

[19] J. Babiak, B. Lucotch, A. Russo, L. Heydt, S. Williams and R. McCaully, J. Automatic Chem., 17 (1995) 55–58.

[20] B.B. Lin, J. Food Drug Anal., 3 (1995) 233–241.

[21] M.C. Needels, D.G. Jones, E.H. Tate, G.L. Heinkel, L.M. Kocher Sperger and W.J. Dower, Proc. Natl. Acad. Sci., 90 (1993) 10700–10704.

[22] R.J. Lipshutz, D. Morris, M. Chee, E. Hubbell, M.J. Kozal, N. Shah, N. Shen, R. Yang and S.P.A. Fodor, Biotechniques, 19 (1995) 442–447.

[23] T. Clackson, H.R. Hoogenboom, A.D. Griffiths and G. Winter, Nature, 352 (1991) 624–628.

[24] H.B. Lowman, S.H. Bass, N. Simpson and J.A. Wells, Biochem., 30 (1991) 10832–10838.

[25] L.J. Garrard, M. Yang, M.P. O'Connell, R.F. Kelley and D.J. Henner, Bio/Technology, 9 (1991) 1373–1377.

[26] W.P.C. Stemmer, Nature, 370 (1994) 389–391.

[27] W.P.C. Stemmer, Proc. Natl. Acad. Sci. USA, 91 (1994) 10747–10751.

[28] R.T. Rowlands, Enz. Micr. Technol., 6 (1984) 3–10.

[29] G. An, J. Bielich, R. Auerbach and E.A. Johnson, Bio/Technology, 9 (1991) 69–73.

[30] C. Ball and M.P. McGonagle, J. Appl. Bacteriol., 45 (1987) 67–74.

[31] R. Spagnoli and L. Cappelletti, Agric. Biol. Chem., 45 (1981) 761–763.

[32] N. Bosworth and P. Towers, Nature, 341 (1989) 167–168.

[33] N. Nelson, Anal. Biochem., 165 (1987) 287–293.

[34] S. Udenfriend, L. Gerber and N. Nelson, Anal. Biochem., 161 (1987) 494–500.

[35] S. Udenfriend, L.D. Gerber, L. Brink and S. Spector, Proc. Natl. Acad. Sci., 82 (1985) 8672–8676.

[36] J.S. Major, J. Receptor Signal Transduction Res., 15 (1995) 595–607.

[37] R. Goodacre, S. Trew, C. Wrigley-Jones, M.J. Neal, J. Maddock, T.W. Ottley, N. Porter and D.B. Kell, Biotechnol. Bioeng., 44 (1994) 1205–1216.

[38] R. Goodacre, S. Trew, C. Wrigley-Jones, G. Saunders, M.J. Neal, N. Porter and D.B. Kell, Anal. Chim. Acta, 313 (1995) 25–43.

[39] R. Goodacre and D.B. Kell, Curr. Opinion Biotechnol., 7 (1996) 20–28.

[40] J.T. Magee, in M. Goodfellow and A.G. O'Donnell (Eds.), Handbook of New Bacterial Systematics, Academic Press, London, 1993, pp. 383–427.

[41] R. Goodacre and D.B. Kell, Anal. Chem., 68 (1996) 271–280.

[42] M.M. Yang and D.C. Youvan, Bio/Technology, 6 (1988) 939–942.

[43] A.P. Arkin, E.R. Goldman, S.J. Robles, C.A. Goddard, W.J. Coleman, M.M. Yang and D.C. Youvan, Bio/Technology, 8 (1990) 746–749.

[44] E.R. Goldman and D.C. Youvan, Bio/Technology, 10 (1992) 1557–1561.

[45] A.P. Arkin and D.C. Youvan, in J. Deisenhofer and J.R. Norris (Eds.), The photosynthetic reaction center, Vol 1., Academic Press, New York, 1993, pp. 133–155.

[46] D. Youvan, Nature, 369 (1994) 79–80.

[47] S. Brown, A.-M. Colson, B. Meunier and P.R. Rich, Eur. J. Biochem., 213 (1993) 137–145.

[48] D. Helm, H. Labischinski and D. Naumann, Microbiol. Methods, 14 (1991) 127–142.

[49] D. Helm, H. Labischinski, G. Schallehn and D. Naumann, J. Gen. Microbiol., 137 (1991) 69–79.

[50] D. Naumann, Infrared Phys., 24 (1983) 233–238.

[51] D. Naumann, V. Fijala and H. Labischinski, Mikrochim. Acta, 1 (1988) 373–377.

[52] D. Naumann, V. Fijjala, H. Lavischinski and P. Giesbrecht, Mol. Struct., 174 (1988) 165–170.

[53] D. Naumann, D. Helm and H. Labischinski, Nature, 351 (1991) 81–82.

[54] D. Naumann, S. Keller, D. Helm, C. Schultz and B. Schrader, J. Mol. Struct., 347 (1995) 399–405.

[55] H.C. Vandermei, D. Naumann and H.J. Busscher, Arch. Oral Biol., 38 (1993) 1013–1019.

[56] P.D. Nichols, J.M. Henson, J.B. Guckert, D.E. Nivens and D.C. White, J. Microbiol. Methods, 4 (1984) 79–94.

[57] D.E. Nivens, J.Q. Chambers, T.R. Anderson, A. Tunlid, J. Smit and D.C. White, J. Microbiol. Methods, 17 (1993) 199–213.

[58] B. Yan, G. Kumaravel, H. Anjaria, A. Wu, R.C. Petter and C.F. Jewell, Jr., J.R. Wareing, J. Org. Chem., 60 (1995) 5736–5738.

[59] H.-U. Gremlich and S.L. Berets, Appl. Spectrosc., 50 (1996) 532–536.

[60] J. Zupan and J. Gasteiger, Neural Networks for Chemists, Verlag Chemie, Weinheim, 1993.

[61] R. Goodacre and R.C.W. Berkeley, FEMS Microbiol. Lett., 71 (1990) 133–138.

[62] S.P. Bouffard, J.E. Katon, A.J. Sommer and N.D. Danielson, Anal. Chem., 66 (1994) 1937–1940.

[63] G. Glauninger, K.A. Kovar and V. Hoffmann, Fresenius J. Anal. Chem., 338 (1990) 710–716.

[64] R. Goodacre, É.M. Timmins, P.J. Rooney, J.J. Rowland and D.B. Kell, FEMS Microbiol. Lett., 140 (1996) 233–239.

[65] H. Wold, in K.R. Krishnaiah, (Ed.), Multivariate Analysis, Academic Press, New York, 1966, pp. 391–420.

[66] H. Hotelling, J. Educational Psychol., 24 (1933) 417–441 and 498–520.

[67] B.S. Everitt, Cluster Analysis, Edward Arnold, London, 1993.

[68] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 1986.

[69] H. Martens and T Næs, Multivariate Calibration, Wiley, Chichester, 1989.

[70] R. Goodacre, Ph.D. Thesis, University of Bristol, Bristol, UK, 1992.

[71] R. Goodacre and D.B. Kell, Anal. Chim. Acta, 279 (1993) 17–26.

[72] R. Goodacre, D.B. Kell and G. Bianchi, J. Sci. Food Agric., 63 (1993) 297–307.

[73] R. Goodacre, A.N. Edmonds and D.B. Kell, J. Anal. Appl. Pyrolysis, 26 (1993) 93–114.

[74] R. Goodacre, M.J. Neal and D.B. Kell, Anal. Chem., 66 (1994) 1070–1085.

[75] R. Goodacre, A. Karim, M.A. Kaderbhai and D.B. Kell, J. Biotechnol., 34 (1994) 185–193.

[76] R. Goodacre, Microbiol. Europe, 2 (1994) 16–22.

[77] R. Goodacre, M.J. Neal, D.B. Kell, L.W. Greenham, W.C. Noble and R.G. Harvey, J. Appl. Bacteriol., 76 (1994) 124–134.

[78] R. Goodacre, Microbiol. Europe, 2 (1994) 16–22.

[79] M.J. Neal, R. Goodacre and D.B. Kell, in Proceedings of the World Congress on Neural Networks, Vol. I, pp. 318–323 International Neural Network Society, San Diego, 1994.

[80] R. Goodacre, S.A. Howell, W.C. Noble and M.J. Neal, Z. Bakteriol., 284 (1996) 501–515.

[81] P.D. Wasserman, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.

[82] D.E. Rumelhart, J.L. McClelland and T.P.R. Group, Parallel Distributed Processing, Experiments in the Microstructure of Cognition, Vols I and II, MIT Press, Cambridge, MA, 1986.

[83] P.J. Werbos, The Roots of Back-propagation: From Ordered Derivatives to Neural Networks and Political Forecasting, Wiley, Chichester, 1994.

[84] J.S. Wong, A.J. Rein, D. Wilks and P. Wilks, Appl. Spectroscopy, 38 (1984) 32–35.

[85] F.L. Bookstein, in J. Kmenta, and J.B. Ramsey (Eds.), Evaluation of Econometric Models, Academic Press, New York, 1980, pp. 75–90.

[86] D.M. Haaland and E.V. Thomas, Anal. Chem., 60 (1988) 1193–1202.

[87] A. Höskuldsson, J. Chemometrics, 2 (1988) 211–228.

[88] A.J. Miller, Subset Selection in Regression, Chapman and Hall, London, 1990.

[89] P.J. Brown, Measurement, Regression and Calibration, Oxford Science Publications, Oxford, 1993.

[90] M.B. Seasholtz and B. Kowalski, Anal. Chim. Acta., 277 (1993) 165–177.

[91] D.B. Kell and B. Sonnleitner, Trends Biotechnol., 13 (1995) 481–492.

[92] P.J. Curran, Prog. Phys. Geography, 18 (1994) 247–266.

[93] W.C. Chiou, Appl. Optics, 24 (1985) 2085–2091.

[94] A.F.H. Goetz, G. Vane, J. Solomon and B.N. Rock, Science, 228 (1985) 1147–1153.

[95] R.J. Birk and T.B. McCord, IEEE Aerospace Electronic Systems Magazine, 9 (1994) 26–33.

[96] D.P. Filiberti, S.E. Marsh and R.A. Schowengerdt, Opt. Eng., 33 (1994) 2520–2528.

[97] G.P. Abousleman, E. Gifford and B.R. Hunt, Opt. Eng., 33 (1994) 2562–2571.

[98] Z.P. Lee, K.L. Carder, S.K. Hawes, R.G. Steward, T.G. Peacock and C.O. Davis, Appl. Optics, 33 (1994) 5721–5732.

[99] S.K. Rogers, J.M. Colombi, C.E. Martin, J.C. Gainey, K. Fielding, T.J. Burns, D.W. Ruck, M. Kabrisky and M. Oxley, Neural Networks, 8 (1995) 1153–1184.

[100] T.A. Wilson, S.K. Rogers and L.R. Myers, Opt. Eng., 34 (1995) 3154–3164.

[101] L.R. Gaddis, L.A. Soderblom, H.H. Kieffer, K.J. Becker, J. Torson and K. Mullins, IEEE Trans. Geosci. Remote Sensing, 34 (1996) 163–178.

[102] J.A. Benediktsson, J.R. Sveinsson and K. Arnason, IEEE Trans. Geosci. Remote Sensing, 33 (1995) 1194–1205.

[103] E. Preston, T. Bergman, R. Gorenflo, D. Hermann, E. Kopala, T. Kuzma, L. Lazofson and R. Orkis, IEEE Aerospace Electronic Systems Magazine, 9 (1994) 13–19.

[104] D. Wienke, W. Vandenbroek, W. Melssen, L. Buydens, R. Feldhoff, T. Kantimm, T. Huthfehre, L. Quick, F. Winter and K. Cammann, Anal. Chim. Acta, 317 (1995) 1–16.

[105] M.J. Persky, Rev. Sci. Instruments, 66 (1995) 4763–4797.

[106] G. Vane, J.E. Duval and J.B. Wellman, in C.M. Pieters and P.A.J. Englert (Eds.), Remote Geochemical Analysis: Elemental and Mineralogical Composition, Cambridge University Press, Cambridge, 1993, pp. 121–143.

[107] J.L. Dwyer, F.A. Kruse and A.B. Lefkoff, Photogrammetric Eng. Remote Sensing, 61 (1995) 1247–1254.

[108] M.P. Fuller and P.R. Griffiths, Anal. Chem., 50 (1978) 1906–1910.

[109] P.R. Griffiths and J.A. de Haseth, Fourier Transform Infrared Spectrometry, Wiley, New York, 1986.

[110] M.B. Mitchell, ACS Adv. Chem. Ser., 236 (1993) 351–375.

[111] B. Schrader, Infrared and Raman Spectroscopy: Methods and Applications, Verlag Chemie, Weinheim, 1995.

[112] B. Hapke, Theory of Reflectance and Emittance Spectroscopy, Cambridge University Press, Cambridge, 1993.